

You will be working with the attached Churn data set available from the UCI Repository of Machine Learning Databases at the University of California, Irvine.

Churn, also called attrition, is a term used to indicate a customer leaving the service of one company in favor of another company. The data set contains 20 predictors worth of information about 3333 customers, along with the target variable, churn, an indication of whether that customer churned (left the company) or not. The variables are as follows:

- State: Categorical, for the 50 states and the District of Columbia
- Account Length: Integer-valued, how long account has been active
- Area code: Categorical
- Phone Number: Essentially a surrogate for customer ID
- International Plan: Dichotomous categorical, yes or no
- Voice Mail Plan: Dichotomous categorical, yes or no
- Number of Voice Mail Messages: Integer-valued
- Total Day Minutes: Continuous, minutes customer used service during the day
- Total Day Calls: Integer-valued
- Total Day Charge: Continuous, perhaps based on above two variables
- Total Eve Minutes: Continuous, minutes customer used service during the evening
- Total Eve Calls: Integer-valued
- Total Eve Charge: Continuous, perhaps based on above two variables
- Total Night Minutes: Continuous, minutes customer used service during the night
- Total Night Calls: Integer-valued
- Total Night Charge: Continuous, perhaps based on above two variables
- Total International Minutes: Continuous, minutes customer used service to make international calls
- Total International Calls: Integer-valued
- Total International Charge: Continuous, perhaps based on above two variables
- Number of Calls to Customer Service: Integer-valued
- Churn: Target. Indicator of whether the customer has left the company (True or False)

Use statistical, predictive and/or visualization techniques in R to analyze the data (for instance, customer usage and satisfaction in comparison to the rate of churning, what attributes might be predictive of churn, why are people leaving, any important correlation, etc.), Present 10 important and interesting findings/predictions with brief explanation of each. Make sure to clean and process your data as appropriate.

You will submit a word document with all your findings and code. You will also present it to me during the specified day and time.